REMARKS

Claims 1-25 are pending and stand rejected. Claims 1, 16, 19, 22, and 23 have been amended.

The specification has been amended to correct typographical errors. No new matter has been added by these amendments.

Applicants respectfully note that the Examiner indicated consideration of the Information Disclosure Statements filed on August 21, 2000, October 23, 2000, February 8, 2001, February 23, 2001, May 20, 2002, November 11, 2002, February 28, 2003, July 28, 2003, and September 11, 2003 but did not indicate consideration of the Information Disclosure Statement filed on November 10, 2003. Applicants respectfully request that the Examiner indicate consideration of the documents submitted with this Information Disclosure Statement by initialing the PTO-1449 form submitted therewith and attaching same to the next communication to Applicants.

The Examiner and the undersigned attorney held a telephone interview on March 23, 2004. The substance of this interview is set forth herein. The Examiner and the undersigned attorney discussed claim 1. The Examiner and the undersigned attorney disagreed about the definition of "maximum quality of service."

Claims 1, 4, 6, 9, and 13-15

Claims 1, 4, 6, 9, and 13-15 stand rejected under 35 U.S.C. § 102(e) as being anticipated by Prieto. Applicants respectfully traverse. As amended, claim 1 recites:

> A method for scheduling a resource to service a plurality of pending requests received from a plurality of schedulable entities, comprising:
>
> selecting a request associated with a schedulable entity, the schedulable entity being associated with a maximum allocation of the resource, the maximum allocation being specified as a maximum quality of service;

responsive to determining that servicing the selected request will exceed the schedulable entity's maximum quality of service, advancing a virtual time for scheduling the requests, without servicing the request; and

responsive to determining that servicing the selected request will not exceed the schedulable entity's maximum quality of service, servicing the request and advancing the virtual time.

The claimed invention relates to a method for scheduling a resource to service a plurality of pending requests received from a plurality of schedulable entities. As recited in claim 1, a request is selected, the request "associated with a schedulable entity, the schedulable entity being associated with a maximum allocation of the resource, the maximum allocation being specified as a maximum quality of service." It is then determined whether "servicing the selected request will exceed the schedulable entity's maximum quality of service." If servicing the request will exceed the maximum quality of service, the request is not serviced. If servicing the request will not exceed the maximum quality of service, the request is serviced. The concept of a maximum quality of service can be used, for example, to prevent a user from receiving more resources than he has paid for, even if this allows resources to be idle during certain times. A resource scheduler that implements a maximum quality of service is often referred to as "non-work-conserving."

Prieto does not disclose, suggest, or teach a maximum allocation of a resource as recited in claim 1. Prieto discusses a protocol that allocates transmission time slots to users assigned to a common resource, namely, a processing satellite communications network (abstract). While Prieto discusses receiving a reservation request from a user and deciding whether to grant, deny, or delay the request (abstract), Prieto does not disclose a user being associated with a "maximum allocation of the resource." Prieto also does not disclose determining whether servicing a request will exceed a user's maximum quality of service and servicing the request (or not) based on this

determination. In direct contrast, Prieto states that the bandwidth (resource allocation) for each user is always at least as big as the user's subscription rate and can be larger if there is excess bandwidth available (10:58-64). Thus, Prieto does not disclose, but instead teaches away from, "the schedulable entity being associated with a maximum allocation of the resource" as recited in claim 1. Accordingly, claim 1 is patentable over Prieto.

Claims 16-18

Claims 16-18[1] stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Prieto, Srinivasan, and Chow in view of Rhee.[2] Applicants respectfully traverse. As amended, claim 16 recites:

> A method for scheduling resource requests from a plurality of schedulable entities, wherein each resource request includes a requested duration and each schedulable entity has a maximum resource allocation, the maximum resource allocation being specified as a maximum quality of service guarantee, the method comprising:
>
> assigning a start number tag to a resource request using a start-time fair queuing algorithm with virtual time scheduling;
>
> selecting the resource request with the smallest start number tag, the selected request having an associated schedulable entity;
>
> limiting the requested duration of the selected resource request to a pre-determined duration upper bound;
>
> servicing the selected resource request if servicing the selected resource request will not exceed the associated schedulable entity's maximum quality of service guarantee; and
>
> advancing a virtual time value.

Prieto does not disclose, suggest, or teach a maximum resource allocation as recited in claim 16. While Prieto discusses receiving a reservation request from a user and deciding whether to grant, deny, or delay the request (abstract), Prieto does not disclose a user being

[1] Although the first sentence of paragraph 6 states that only claim 16 is rejected, the remainder of the paragraph also discusses the rejections of claims 17 and 18.

associated with a "maximum resource allocation." Prieto also does not disclose "servicing the selected resource request if servicing the selected resource request will not exceed the associated schedulable entity's maximum quality of service guarantee." In direct contrast, Prieto states that the bandwidth (resource allocation) for each user is always at least as big as the user's subscription rate and can be larger if there is excess bandwidth available (10:58-64). Thus, Prieto does not disclose, but instead teaches away from, "each schedulable entity has a maximum resource allocation" as recited in claim 16.

Srinivasan does not remedy this deficiency. Srinivasan discusses a queue selection method for performing data transfers (abstract). While Srinivasan mentions passing packets to queues based on the packets' bandwidth requirements and the queues' bandwidth assignments (4:4:45-46), Srinivasan does not disclose a packet or flow of packets having a maximum resource allocation. In direct contrast, a bandwidth assignment for a queue in Srinivasan represents a minimum flow rate (6:12-14). Srinivasan also does not disclose "servicing the selected resource request if servicing the selected resource request will not exceed the associated schedulable entity's maximum quality of service guarantee." Thus, Srinivasan does not disclose, but instead teaches away from, "each schedulable entity has a maximum resource allocation" as recited in claim 16.

Chow also does not remedy this deficiency. Chow discusses a method for servicing queues that hold messages, such as data packets, for subsequent processing or transmission (abstract). While Chow discusses provisioning each queue with a minimum guaranteed service rate (3:21-23), Chow does not disclose a queue having a maximum resource allocation. Chow

[2] Although the first sentence of paragraph 6 cites "Rahee'008", Applicants respectfully note that the first-named inventor of the 6,457,008 patent is "Rhee" as stated in paragraph 3.

also does not disclose "servicing the selected resource request if servicing the selected resource request will not exceed the associated schedulable entity's maximum quality of service guarantee." In direct contrast, the hierarchical scheduler in Chow comprises a work-conserving idle bandwidth scheduler. Thus, idle bandwidth will be distributed among the queues. Thus, Chow does not disclose, but instead teaches away from, "each schedulable entity has a maximum resource allocation" as recited in claim 16.

Rhee also does not remedy this deficiency. Rhee discusses a resource scheduler for allocating a computer system resource among a set of processes (abstract). While Rhee discusses allocating a resource among classes of requesting processes according to dynamically selectable scheduling plans (3:13-15), Rhee does not disclose a process or class of processes having a maximum resource allocation. Rhee also does not disclose "servicing the selected resource request if servicing the selected resource request will not exceed the associated schedulable entity's maximum quality of service guarantee." In direct contrast, the scheduler in Rhee allows a process to continue executing, even if it has used its specified resource allocation, as long as no other processes are waiting to use the resource (11:57-12:13). Thus, Rhee does not disclose "each schedulable entity has a maximum resource allocation" as recited in claim 16.

Since Prieto, Srinivasan, Chow, and Rhee do not individually disclose "each schedulable entity has a maximum resource allocation," it follows that the combination of these references cannot disclose or suggest this claimed element. Accordingly, claim 16 is patentable over Prieto, Srinivasan, Chow, and Rhee, both individually and in combination. For the record, Applicants do not agree with the Examiner that there exists motivation to combine these four references.

Claims 19-20

Claims 19-20 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Prieto in view of Beshai.[3] Applicants respectfully traverse. As amended, claim 19 recites:

> A system for scheduling pending resource requests from a plurality of schedulable entities while limiting a maximum resource allocation, the maximum resource allocation being specified as a maximum quality of service allocated to each schedulable entity, comprising:
>
> a plurality of schedulable entity queues for holding pending resource requests, each schedulable entity queue holding resource requests from a schedulable entity;
>
> a scheduler for selecting resource requests from the plurality of schedulable entity queues using a fair-share scheduling algorithm, and further adapted to increment a virtual time value each time a resource request is selected; and
>
> a plurality of rate controllers associated with the plurality of schedulable entity queues, each rate controller adapted to limit the rate at which resource requests selected by the scheduler are serviced to the schedulable entity's maximum quality of service.

Prieto does not disclose, suggest, or teach a maximum resource allocation as recited in claim 19. While Prieto discusses receiving a reservation request from a user and deciding whether to grant, deny, or delay the request (abstract), Prieto does not disclose a user being associated with a "maximum resource allocation." Prieto also does not disclose "each rate controller adapted to limit the rate at which resource requests selected by the scheduler are serviced to the schedulable entity's maximum quality of service." In direct contrast, Prieto states that the bandwidth (resource allocation) for each user is always at least as big as the user's subscription rate and can be larger if there is excess bandwidth available (10:58-64). Thus, Prieto does not disclose, but instead teaches away from, "the maximum resource allocation being

[3] Although the first sentence of paragraph 7 cites "Beshi'721", Applicants respectfully note that the first-named inventor of the 6,580,721 patent is "Beshai" as stated in the Notice of References Cited.

specified as a maximum quality of service allocated to each schedulable entity" as recited in claim 19.

Beshai does not remedy this deficiency. Beshai discusses a method and network for transferring data packets at a regulated bit rate (abstract). While Beshai discusses a packet having a QOS (quality of service) index, this index indicates a rank of the packet with respect to other data packets (10:18-32). In particular, the QOS index does not indicate capacity allocation. (*Id.*) Thus, Beshai does not disclose, but instead teaches away from, a packet being associated with a "maximum resource allocation." Beshai also does not disclose "each rate controller adapted to limit the rate at which resource requests selected by the scheduler are serviced to the schedulable entity's maximum quality of service." Thus, Beshai does not disclose "the maximum resource allocation being specified as a maximum quality of service allocated to each schedulable entity" as recited in claim 19.

Since Prieto and Beshai do not individually disclose "the maximum resource allocation being specified as a maximum quality of service allocated to each schedulable entity," it follows that the combination of these references cannot disclose or suggest this claimed element. Accordingly, claim 19 is patentable over Prieto and Beshai, both individually and in combination. For the record, Applicants do not agree with the Examiner that there exists motivation to combine these two references.

Claim 22

Claim 22 stands rejected under 35 U.S.C. § 103(a) as being unpatentable over Prieto and Beshai[4] in view of Ganmukhi. Applicants respectfully traverse. As amended, claim 22 recites:

[4] Although the first sentence of paragraph 9 cites "Beshi'721", Applicants respectfully note that the first-named inventor of the 6,580,721 patent is "Beshai" as stated in the Notice of References Cited.

A hierarchical system for scheduling resource requests from a plurality of child schedulable entities while limiting a maximum resource allocation allocated to a plurality of parent schedulable entities, the maximum resource allocation being specified as a maximum quality of service, comprising:

a plurality of child schedulable entity queues for holding pending resource requests, each child schedulable entity queue holding resource requests from a child schedulable entity;

one or more child schedulers for selecting resource requests from the plurality of child schedulable entity queues using a fair-share scheduling algorithm, and further adapted to transmit selected resource requests to a parent schedulable entity queue;

a plurality of parent schedulable entity queues, each parent schedulable entity queue receiving resource requests from a subset of the child schedulable entity queues, each parent schedulable entity queue holding resource requests received from one of the child schedulers;

a parent scheduler for selecting resource requests from the plurality of parent schedulable entity queues using a fair-share scheduling algorithm, and further adapted to increment a virtual time value each time a resource request is selected; and

a plurality of rate controllers associated with the plurality of parent schedulable entity queues, each rate controller adapted to limit the rate at which resource requests selected by the parent scheduler are serviced to a parent schedulable entity's maximum quality of service.

Prieto does not disclose, suggest, or teach a maximum resource allocation as recited in claim 22. While Prieto discusses receiving a reservation request from a user and deciding whether to grant, deny, or delay the request (abstract), Prieto does not disclose a user being associated with a "maximum resource allocation." Prieto also does not disclose "each rate controller adapted to limit the rate at which resource requests selected by the parent scheduler are serviced to a parent schedulable entity's maximum quality of service." In direct contrast, Prieto states that the bandwidth (resource allocation) for each user is always at least as big as the user's subscription rate and can be larger if there is excess bandwidth available (10:58-64). Thus, Prieto

does not disclose, but instead teaches away from, "a maximum resource allocation allocated to a plurality of parent schedulable entities" as recited in claim 22.

Beshai does not remedy this deficiency. Beshai discusses a method and network for transferring data packets at a regulated bit rate (abstract). While Beshai discusses a packet having a QOS (quality of service) index, this index indicates a rank of the packet with respect to other data packets (10:18-32). In particular, the QOS index does not indicate capacity allocation. (*Id.*) Thus, Beshai does not disclose, but instead teaches away from, a packet being associated with a "maximum resource allocation." Beshai also does not disclose "each rate controller adapted to limit the rate at which resource requests selected by the parent scheduler are serviced to a parent schedulable entity's maximum quality of service." Thus, Beshai does not disclose "a maximum resource allocation allocated to a plurality of parent schedulable entities" as recited in claim 22.

Ganmukhi also does not remedy this deficiency. Ganmukhi discusses a method of fairly and efficiently scheduling transmissions of a packet from a plurality of sessions onto a network. While Ganmukhi discusses an input session having a QOS (quality of service), this QOS does not indicate a maximum resource allocation (3:5-7; 4:8-14). In particular, none of the six QOS traffic classes indicates a maximum resource allocation (4:15-52). Thus, Ganmukhi does not disclose "a maximum resource allocation allocated to a plurality of parent schedulable entities" as recited in claim 22.

Since Prieto, Beshai, and Ganmukhi do not individually disclose "a maximum resource allocation allocated to a plurality of parent schedulable entities," it follows that the combination of these references cannot disclose or suggest this claimed element. Accordingly, claim 22 is patentable over Prieto, Beshai, and Ganmukhi, both individually and in combination. For the

record, Applicants do not agree with the Examiner that there exists motivation to combine these three references.

Claim 23

Claim 23 stands rejected under 35 U.S.C. § 103(a) as being unpatentable over Prieto in view of Srinivasan. Applicants respectfully traverse. As amended, claim 23 recites:

> A computer program product for scheduling a plurality of pending requests for service from a resource received from a plurality of schedulable entities, while preventing each schedulable entity from receiving an amount of the resource that exceeds a maximum resource allocation allocated to each schedulable entity, the maximum resource allocation being specified as a maximum quality of service, the computer program product comprising a computer readable medium that stores program code, including
>
> program code that selects a request associated with a schedulable entity using a fair-share scheduling algorithm;
>
> program code that services the request if a rate controller determines that servicing the request will not exceed the associated schedulable entity's maximum quality of service; and
>
> program code that advances a virtual time in the fair-share scheduling algorithm.

Prieto does not disclose, suggest, or teach a maximum resource allocation as recited in claim 23. While Prieto discusses receiving a reservation request from a user and deciding whether to grant, deny, or delay the request (abstract), Prieto does not disclose a user being allocated a "maximum resource allocation." Prieto also does not disclose "servic[ing] the request if a rate controller determines that servicing the request will not exceed the associated schedulable entity's maximum quality of service." In direct contrast, Prieto states that the bandwidth (resource allocation) for each user is always at least as big as the user's subscription rate and can be larger if there is excess bandwidth available (10:58-64). Thus, Prieto does not

disclose, but instead teaches away from, "a maximum resource allocation allocated to each schedulable entity" as recited in claim 23.

Srinivasan does not remedy this deficiency. Srinivasan discusses a queue selection method for performing data transfers (abstract). While Srinivasan mentions passing packets to queues based on the packets' bandwidth requirements and the queues' bandwidth assignments (4:4:45-46), Srinivasan does not disclose a packet or flow of packets having a maximum resource allocation. In direct contrast, a bandwidth assignment for a queue in Srinivasan represents a minimum flow rate (6:12-14). Srinivasan also does not disclose "servic[ing] the request if a rate controller determines that servicing the request will not exceed the associated schedulable entity's maximum quality of service." Thus, Srinivasan does not disclose, but instead teaches away from, "a maximum resource allocation allocated to each schedulable entity" as recited in claim 23.

Since Prieto and Srinivasan do not individually disclose "a maximum resource allocation allocated to each schedulable entity," it follows that the combination of these references cannot disclose or suggest this claimed element. Accordingly, claim 23 is patentable over Prieto and Srinivasan, both individually and in combination. For the record, Applicants do not agree with the Examiner that there exists motivation to combine these two references.

The claims not specifically mentioned above incorporate the features of their respective base claims and are patentable for at least the same reasons.

Applicants respectfully submit that the pending claims are now allowable over the cited art of record and request that the Examiner allow this case. The Examiner is invited to contact the undersigned in order to advance the prosecution of this application.

Respectfully submitted,
PAWAN GOYAL, ET AL.

By: Sabra-Anne Truesdale 6-17-04
Sabra-Anne R. Truesdale
Reg. No. 55,687
Fenwick & West LLP
Silicon Valley Center
801 California Street
Mountain View, CA 94041
Tel.: (650) 335-7187
Fax.: (650) 938-5200